The Korean Communications
in Statistics Vol. 8, No. 2, 2001
pp. 327-336

Efficiency of Aggregate Data in Non-linear Regression[1]

Jib Huh[2]

Abstract

This work concerns estimating a regression function, which is not linear, using aggregate data. In much of the empirical research, data are aggregated for various reasons before statistical analysis. In a traditional parametric approach, a linear estimation of the non-linear function with aggregate data can result in unstable estimators of the parameters. More serious consequence is the bias in the estimation of the non-linear function. The approach we employ is the kernel regression smoothing. We describe the conditions when the aggregate data can be used to estimate the regression function efficiently. Numerical examples will illustrate our findings.

Keywords : Local linear fitting, Within group variance, Mean integrated squared error

1. Introduction

Sometimes, only aggregate data summarized individual data in each groups or areas have been observed, and individual data are unavailable. An aggregate data analysis is based not on the individual observations but instead on their averages. Consider the bivariate full data $(X_{i1}, Y_{i1}), \cdots, (X_{in_i}, Y_{in_i})$ which form an independent and identically distributed sample from a population $(X_i, Y_i), i=1, \cdots, M$. Of interest is to estimate the regression function $m(x) = E(Y_i | X_i = x)$. Especially, we are interested in estimating the regression function $m(x)$ based on the observed sample $\{(\overline{X}_i, \overline{Y}_i), i=1, \cdots, M\}$ where $\overline{X}_i = \sum_{j=1}^{n_i} X_{ij}/n_i$ and $\overline{Y}_i = \sum_{j=1}^{n_i} Y_{ij}/n_i$.

When a regression function was linear, Aitkin and Longford (1986) estimated the regression model based on aggregate data unbiasedly, and showed the estimator would be highly

1) This research was supported by a Full-time Fellowship Award from the Alberta Heritage Foundation for Medical Research (AHFMR), Canada, and the Brain Korea 21, Korea.
2) Postdoctral Fellow, Department of Statistics, Seoul National University, Seoul 151-747
E-mail : huhjib@stats.snu.ac.kr

unstable with a very large standard error due to the sampling variation. Prentice and Sheppard (1995) introduced a method for the analysis of aggregate data studies, with binary responses. Carroll (1997) studied the effect of measurement error on the Pretice and Sheppard unweighted aggregate data analysis with the various populations.

When the regression function is non-linear, we may have serious biased estimators for the non-linear function. This is illustrated in Figure 1. The solid curve is the true regression function. Assume that three individual data points represented by dots are observed in a group, and we get an averaged data point represented by an asterisk. The aggregate data point seems to be far from the true value of the function.

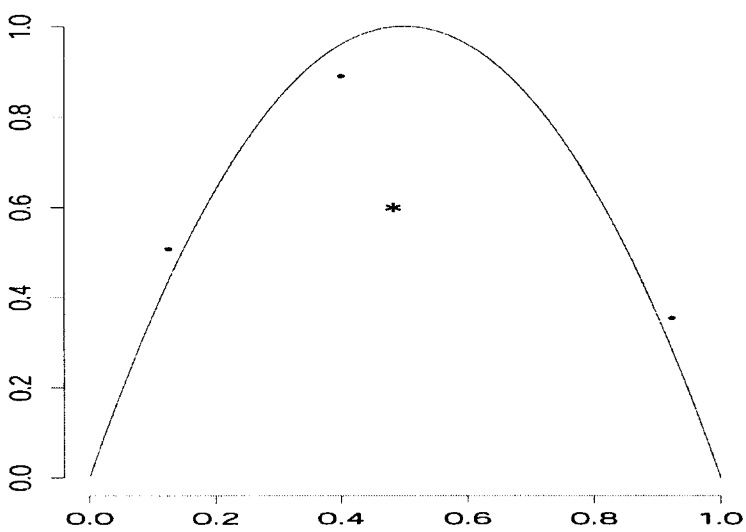

Figure 1: True regression function (solid curve), three individual data points (represented by dots) and aggregate data point (represented by an asterisk)

We will explain intuitively why estimators for the regression function based on aggregate data has a large bias. It is enough to consider that the regression function is quadratic,

$$Y_{ij} = \beta_0 + \beta_1 X_{ij} + \beta_2 X^2_{ij} + \varepsilon_{ij}, \quad i = 1, \cdots, M; \ j = 1, \cdots, n_i \ , \tag{1}$$

where the error ε_{ij} are independent and identically distributed random variables with mean 0 and variance $\sigma^2 < \infty$. Then, the regression model based on $\{(\overline{X}_i, \overline{Y}_i), i = 1, \cdots, M\}$ could be

$$\overline{Y}_i = \beta_0 + \beta_1 \overline{X}_i + \beta_2 \overline{X}_i^2 + \varepsilon^*_i, \quad i = 1, \cdots, M. \tag{2}$$

Here the errors ε^*_i, different from ε_{ij}, are independent and identically distributed random variables with mean 0 and variance σ^2 / n_i. However, the true regression function (1) averaged by index j can be written as

$$\overline{Y_i} = \beta_0 + \beta_1 \overline{X_i} + \beta_2 \sum_{j=1}^{n_i} X_{ij}^2 / n_i + \overline{\varepsilon_i}, \qquad i = 1, \cdots, M, \tag{3}$$

where $\overline{\varepsilon_i} = \sum_{j=1}^{n_i} \varepsilon_{ij}/n_i$, $i = 1, \cdots, M$. According to the difference between the conditional mean parts in (2) and (3), the model (2) has a model bias related to the within group variances $\beta_2 (\sum_{j=1}^{n_i} X_{ij}^2 / n_i - \overline{X_i}^2)$. The assumption that the errors ε_i^* have mean 0 makes the model bias. In order that the model (2) may be equal to the true regression model (1), the errors ε_i^* must be assumed that the conditional means of them are $\beta_2 (\sum_{j=1}^{n_i} X_{ij}^2 / n_i - \overline{X_i}^2)$.

In this paper, we discuss the conditions when the aggregate data can be suitable to estimate the non-linear regression function without assuming a specific parametric form. In Section 2, the asymptotic bias and variance of the kernel type nonparametric estimator for the regression are described. Section 3 shows simulation results. Finally some technical arguments are deferred in Section 4.

2. Local Modeling and Estimation

The usual nonparametric assumption is that the full data satisfy the relationship

$$Y_{ij} = m(X_{ij}) + \varepsilon_{ij}, \qquad i = 1, \cdots, M;\ j = 1, \cdots, n_i \tag{4}$$

where the errors ε_{ij} are independent random variables with mean zero and variance $\sigma^2 < \infty$. We suppose for the sake of definiteness that the regression function m is defined on $[0,1]$. There now exist many methods for obtaining a nonparametric regression estimate for m. Some widely used methods are those based on kernel functions, spline functions, and wavelets. Each of these methods has its own merits. The kernel based approach is, however, preferable because of its mathematical and intuitive simplicity. Stone (1977), Cleveland (1979), Müller (1987) and Fan (1992) studied a class of kernel type regression estimators called local polynomial estimators. Local polynomial estimator is more attractive method both from theoretical and practical point of view than other commonly used kernel type estimators. See Wand and Jones (1995), and Fan and Gijbels (1996). For simplicity, we choose local linear estimator for our study.

To estimate the regression function based on the aggregate data $\{(\overline{X_i}, \overline{Y_i}), i = 1, \cdots, M\}$, we approximate the unknown regression function m locally by a linear function. Let K be a kernel function and h be a bandwidth. At a point x an estimator for $m(x)$ is obtained by fitting the linear function $\alpha + \beta(\cdot - x)$ to the $(\overline{X_i}, \overline{Y_i})$ using weighted least squares with kernel weights $K_h(\overline{X_i} - x)$: minimize

$$\sum_{i=1}^{M} \{ \overline{Y_i} - \alpha - \beta(\overline{X_i} - x) \}^2 K_h(\overline{X_i} - x) \tag{5}$$

where $K_h(u) = h^{-1}K(u/h)$. Denote by $\hat{\alpha}$ and $\hat{\beta}$ the minimizers of (5). We suggest that an estimator for $m(x)$ is

$$\widehat{m}(x) = \hat{\alpha}. \tag{6}$$

We consider that the number of data in each group n_i is fixed.

Before describing the asymptotic bias and variance, we need some assumptions. For simplicity, the number of data n_i in the ith group is fixed as n, for all i. Let $N = M \cdot n$ denote the number of the full data and g_i denote the design density of X_i. Define $[a_i, b_i]$ as the support of the design density g_i. Let us assume that the function m is twice continuously differentiable on $[0,1]$ and the bandwidth h satisfies $h \to 0$ and $Mh \to \infty$ as $M \to \infty$. The latter assumption implies that $Nh \to \infty$ as $N \to \infty$ since the number of data in each group n is fixed. Further, assume that the lengths of the supports $L_i \equiv b_i - a_i$ go to zero faster than the bandwidth h does, as $M \to \infty$ and $\cup_{i=1}^{M}[a_i, b_i] = [0,1]$, for every M. This assumption means that the interval $[0,1]$ can be finely divided as increasing the number of groups M and the union of supports is $[0,1]$. Define $\bar{f}(x) = \lim_{M\to\infty} \sum_{i=1}^{M} f_i(x)/M$ where the function f_i is the density of $\overline{X_i}$ which is not degenerated since the number of data n is fixed. We assume that the function $\bar{f}$ is bounded away from zero on $[0,1]$. Suppose that the kernel K is symmetric about zero and a probability density with support $[-1,1]$.

Write $K_h^2(u) = K^2(u/h)/h$ and $SSW_i = \sum_{j=1}^{n}(X_{ij} - \overline{X_i})^2/n$. Let

$$\overline{S_l} = \sum_{i=1}^{M}(\overline{X_i} - x)^l K_h(\overline{X_i} - x)/M,$$

$$\overline{T_l} = \sum_{i=1}^{M} SSW_i(\overline{X_i} - x)^l K_h(\overline{X_i} - x)/M,$$

$$\overline{U_l} = \sum_{i=1}^{M}(\overline{X_i} - x)^l K_h^2(\overline{X_i} - x)/M,$$

where $l = 0,1,2,3$. Under the assumptions described above, the expectation and variance of the estimator $\widehat{m}(x)$ in (6) is given by

$$\text{bias}(\widehat{m}(x)|\overline{X_1}, \cdots, \overline{X_M}) = \frac{m''(x)}{2}\frac{1}{\overline{S_0}\,\overline{S_2} - \overline{S_1}^2}(\overline{S_2}^2 - \overline{S_1}\,\overline{S_3} + \overline{S_2}\,\overline{T_0} - \overline{S_1}\,\overline{T_1})(1 + o_p(1)), \tag{7}$$

$$\text{var}(\widehat{m}(x)|\overline{X_1}, \cdots, \overline{X_M}) = \frac{\sigma^2}{Nh}\frac{1}{(\overline{S_0}\,\overline{S_2} - \overline{S_1}^2)^2}(\overline{S_2}^2\overline{U_0} - 2\overline{S_1}\,\overline{S_2}\,\overline{U_1} + \overline{S_1}^2\overline{U_2}).$$

The kernel weighted local linear fitting makes the first two terms within the brackets of the bias in (7). On the other hand, the bias has the extra terms kernel weighted the within group variances SSW_i at center x. Therefore, the asymptotic bias depends how large the within group variances are. It means that the bias can be reduced if we can divide finely the domain of m to get small the within group variances. However, the variance term does not depend on the within group variances. Now, we will show that the rates of convergences of the asymptotic bias and variance are same with those based on the full data. For a unifying treatment of interior and boundary points x, we consider an effective range of integration. Let $I \equiv I(x,h) = [(x-1)/h, x/h] \cap [-1,1]$, $\nu_j \equiv \nu_j(I) = \int_I z^j K(z)dz$ and $\omega_j \equiv \omega_j(I) = \int_I z^j K^2(z)dz$. As h goes to zero, the interval $I(x,h)$ reduces to $[-1,1]$ for an interior point $x \in [h, 1-h]$, and to $[-1, c_1]$ or $[-c_2, 1]$ for a boundary point of the form $c_1 h$ or $1-c_2 h$ $(0 \le c_1, c_2 < 1)$. To approximate the leading bias and variance in (7) analyses similar to those used in Wand and Jones (1995, see page 123-124) give

$$\overline{S}_l = h^l \nu_l(K) \overline{f}(x)(1+o_p(1)) \text{ and } \overline{U}_l = h^l \omega_l \overline{f}(x)(1+o_p(1)),$$

for $l = 0,1,2,3$. Let $K_1(u) = (\nu_0\nu_2 - \nu_1)^{-1}(\nu_2 - u\nu_1)K(u)$. Write $x_1 = \int_I u^2 K_1(u)du$, and $x_2 = \int_I \{K_1(u)\}^2 du$. Now, the estimator $\widehat{m}(x)$ has pointwise asymptotic bias and variance,

$$\text{asymp.bias}\{\widehat{m}(x)\} = \frac{1}{2} x_1 m''(x)h^2 \quad \text{and} \quad \text{asymp.var}\{\widehat{m}(x)\} = \frac{\sigma^2}{Nh} x_2 \frac{1}{\overline{f}}(x). \tag{8}$$

The asymptotic bias in (8) is same with the leading bias term of the local linear regression estimator based on the full data. On the other hand, the asymptotic variance is almost unchanged. Its denominator has the limit value of the average densities at the point x. Härdle and Grund (1991), Härdle and Scott (1992), and Fan and Marron (1994) have studied the regression function estimation based on the binned data to get fast computing algorithms. The main philosophy of their proposed methods is to use a small data set summarized the binned data instead of a full data set. We also reduced the computing time with the aggregated data. Hall, Park and Turlach (1998) have considered the regression function estimation based on the transformed and binned data to overcome problems associated with irregularly-spaced design. In the case of the bin width going to zero faster than the bandwidth, they showed that the rate of convergence of their estimator was same with that of the estimator using the full data set. In our case, since the lengths of the supports of the design densities are decreasing as increasing the number of groups M and it converges to zero faster than the bandwidth, the last two terms within the brackets in (7) converges to zero faster than the first two terms. Then, the estimator $\widehat{m}$ is a consistent estimator and has the same rate of convergence of the local linear estimator based on the full data set. The difference between binned data and aggregate data is that all of bins are disjoint but the interval in which data are aggregated does not need to be disjoint.

3. Numerical Experiments

To investigate the performance of the aggregate data, a simulation study is carried out. Consider the model

$$m(x) = 4(x-0.5) + \frac{3}{2}\exp(-64(x-0.5)^2)I(0 \le x \le 1). \tag{9}$$

For the calculation of the estimations, we use the Epanechnikov kernel function

$$K(x)=\frac{3}{4}(1-x^2)I(|x|\le 1).$$

We try several values of M and n. We consider three and four types of M and n, respectively, such as M=30, 50, 100 and n=5, 10, 50, 100. Furthermore, we choose two kinds of distributions of X_i. In the first case, the covariate X_{ij} is distributed as

$$X_{ij}\sim \text{Uniform}\left(\frac{[(i-1)/10]}{M/10}, \frac{[((i-1)/10)+1]}{M/10}\right), \; i=1,\cdots,M; j=1,\cdots,n,$$

where $[\ell]$ is the largest integer that does not exceed ℓ. Next, we select the beta distribution $\text{Beta}(\alpha, 10-\alpha)$. When M=30, 50, and 100, the parameter α of the distribution are $2.5\times[((i-1)/10)+1]$, $2\times[((i-1)/10)+1]-1$, and $[((i-1)/10)+1]-0.5$, respectively. In the second case, all of the beta distributions have the same support $[0,1]$ on which the regression function m in (9) is defined. We can guess that the local linear estimator for the regression function based on aggregate data under these distributions must have the poor performance since the support $[0,1]$ is not finely divided as increasing M as we assumed. On the other hand, in the first case, the support $[0,1]$ is divided several parts of which lengths are decreased as increasing M. We compare the performance of local linear estimator under the first design case with that under the second design case. The Gaussian white noise with $\sigma=0.5$ is added to produce the simulated data. Table 1 and 2 contain the results of the simulation based on 1000 pseudo samples. We compute the Monte Carlo estimates of the integrated mean square errors (MISE) for various bandwidths h, but the minimum MISE's are reported in the tables with the optimal bandwidths. We also give the Monte Carlo estimates of the integrated squared biases (IBIAS) and variances (IVAR) corresponding to the minimum MISE's. In each every two rows, the first one is the results based on the full data, and the second one is those based on the aggregate data.

Table 1: The Monte Carlo estimates IBIAS, IVAR and MISE with standard errors in brackets based on the full data and the aggregate data when the design densities are uniform distributions.

(M,n)	h	IBIAS	IVAR	MISE
(30,5)	0.10	0.003491	0.014154 (0.000228)	0.017645 (0.000268)
	0.20	0.046165	0.025688 (0.000707)	0.071853 (0.000785)
(30,10)	0.08	0.001595	0.008121 (0.000113)	0.009716 (0.000126)
	0.22	0.043639	0.019898 (0.000566)	0.063538 (0.000614)
(30,50)	0.06	0.000516	0.001959 (0.000024)	0.002475 (0.000029)
	0.22	0.039808	0.018459 (0.000661)	0.058267 (0.000684)
(30,100)	0.05	0.000268	0.001123 (0.000012)	0.001392 (0.000015)
	0.23	0.040654	0.017878 (0.000529)	0.058532 (0.000547)

(Table 1 continued)

(M, n)	h	IBIAS	IVAR	MISE
(50,5)	0.09	0.002461	0.008876 (0.000130)	0.011338 (0.000159)
	0.14	0.013732	0.014867 (0.000361)	0.028599 (0.000407)
(50,10)	0.07	0.000896	0.005265 (0.000066)	0.006161 (0.000074)
	0.14	0.012977	0.012182 (0.000303)	0.025159 (0.000325)
(50,50)	0.05	0.000255	0.001389 (0.000015)	0.001644 (0.000018)
	0.14	0.011679	0.011195 (0.000354)	0.022891 (0.000356)
(50,100)	0.04	0.000116	0.000830 (0.000008)	0.000947 (0.000009)
	0.14	0.011386	0.011527 (0.000323)	0.022913 (0.000325)
(100,5)	0.07	0.000924	0.005199 (0.000068)	0.006123 (0.000076)
	0.08	0.003920	0.008198 (0.000165)	0.012118 (0.000174)
(100,10)	0.06	0.000521	0.002909 (0.000033)	0.003430 (0.000040)
	0.08	0.004433	0.006156 (0.000153)	0.010589 (0.000160)
(100,50)	0.04	0.000115	0.000839 (0.000008)	0.000954 (0.000010)
	0.09	0.005517	0.003810 (0.000133)	0.009327 (0.000138)
(100,100)	0.04	0.000109	0.000418 (0.000004)	0.000527 (0.000005)
	0.10	0.006022	0.003361 (0.000116)	0.009383 (0.000138)

Table 2: The Monte Carlo estimates IBIAS, IVAR and MISE with standard errors in brackets based on the full data and the aggregate data when the design densities are beta distributions.

(M, n)	h	IBIAS	IVAR	MISE
(30,5)	0.15	0.013594	0.024691 (0.000845)	0.038285 (0.000865)
	0.30	0.121467	0.046456 (0.001712)	0.167923 (0.002146)
(30,10)	0.12	0.006529	0.015110 (0.000386)	0.021639 (0.000391)
	0.29	0.119791	0.038193 (0.001345)	0.157984 (0.001672)
(30,50)	0.08	0.001600	0.004507 (0.000104)	0.006107 (0.000109)
	0.27	0.113669	0.030939 (0.000952)	0.144608 (0.001287)
(30,100)	0.07	0.001043	0.002568 (0.000058)	0.003611 (0.000062)
	0.27	0.112624	0.028251 (0.001035)	0.140875 (0.001366)
(50,5)	0.09	0.002392	0.008007 (0.000112)	0.010399 (0.000136)
	0.13	0.077055	0.013191 (0.000321)	0.090246 (0.000707)
(50,10)	0.07	0.000974	0.004957 (0.000058)	0.005932 (0.000071)
	0.13	0.083272	0.008840 (0.000235)	0.092112 (0.000541)
(50,50)	0.05	0.000279	0.001357 (0.000015)	0.001636 (0.000018)
	0.11	0.081878	0.008152 (0.000200)	0.090030 (0.000341)
(50,100)	0.05	0.000263	0.000667 (0.000007)	0.000930 (0.000010)
	0.11	0.081141	0.007509 (0.000183)	0.088651 (0.000263)
(100,5)	0.07	0.000930	0.004792 (0.000061)	0.005722 (0.000072)
	0.08	0.072118	0.006855 (0.000097)	0.078973 (0.000528)
(100,10)	0.06	0.000528	0.002828 (0.000033)	0.003356 (0.000039)
	0.07	0.077638	0.004464 (0.000063)	0.082102 (0.000398)
(100,50)	0.05	0.000269	0.000648 (0.000007)	0.000917 (0.000010)
	0.06	0.081665	0.002075 (0.000045)	0.083739 (0.000186)
(100,100)	0.04	0.000113	0.000399 (0.000004)	0.000513 (0.000005)
	0.06	0.082260	0.001568 (0.000037)	0.083823 (0.000132)

In the tables, IBIAS's and optimal bandwidths based on the aggregate data are not changed quietly as increasing n in each case M. But, IVAR's are inclined to be decreased. The minimum MISE's of the estimators based on aggregate data under the first design cases are smaller than those under the second design cases, due to smaller within group variances we described in Section 2. This means that the performances of local linear estimators depend on the lengths of supports of the design densities.

Figure 2 depicts, for a simulated dataset from the uniform and the beta distribution in the case $M, n = 100$, the true regression function denoted by solid curve, and the estimated regression functions under the uniform and the beta design cases represented by dashes. To see that the aggregate date are far from the true regression function, we plot the aggregate data in this figure. The true regression function has non-linear curvature around 0.5 and 0.7. The estimator under the beta distributions has large bias near the points.

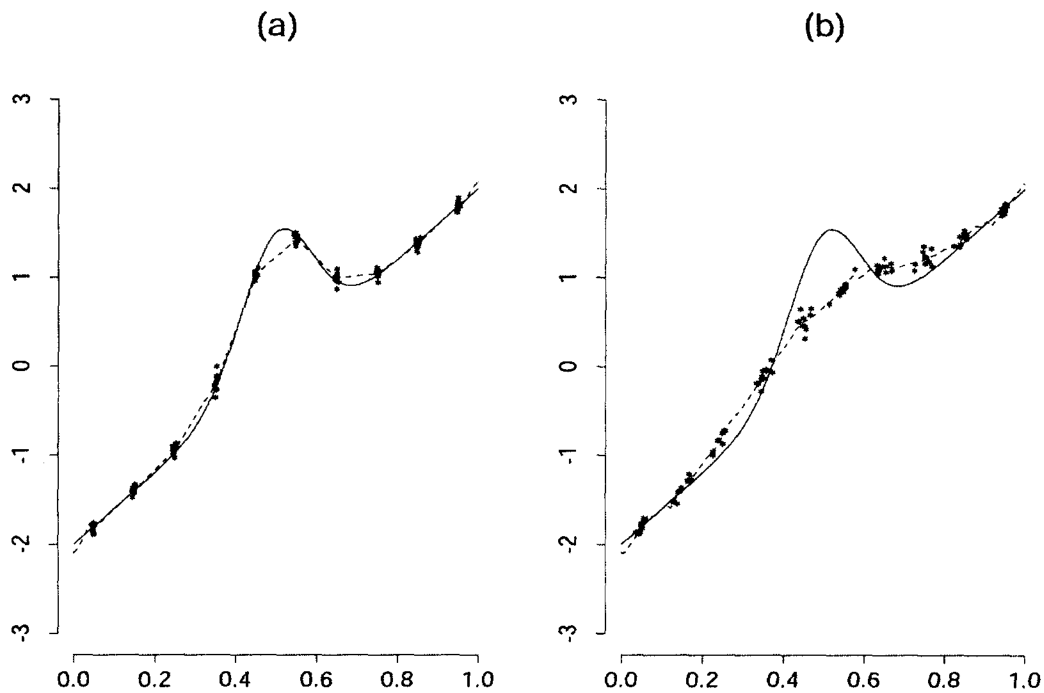

Figure 2: True regression function (solid curve) and regression function estimates (dashed curves) based on aggregate data (asterisks) under (a) the uniform design density and (b) the beta design density.

4. Technical Arguments

We will show some technical arguments for (8) here. Define $\overline{X}$ be $M \times 2$ matrix having their (i,j)th elements equal to $(\overline{X_i}-x)^{j-1}$, and $\overline{Y}=(\overline{Y_1},\cdots,\overline{Y_M})$. Further, let $\overline{W}$ be an $M\times M$ diagonal matrix of weights having the (i,i)th entry equal to $K_h(\overline{X_i}-x)$. Assuming invertibility $\overline{X}^T\overline{W}\overline{X}$, the standard weighted least squares theory lead to the solution of (6) $\hat{a}=(\overline{X}^T\overline{W}\overline{X})^{-1}\overline{X}^T\overline{W}\overline{Y}$. Let $B=(\sum_{j=1}^{n} m(X_{1j})/n,\cdots,\sum_{j=1}^{n} m(X_{Mj})/n)^T$. Then, the conditional expectation and variance of the estimator (6) can be written as follows:

$$
\begin{aligned}
E(\hat{m}(x)|\overline{X_1},\cdots,\overline{X_M}) &= e_1^T(\overline{X}^T\overline{W}\overline{X})^{-1}\overline{X}^T\overline{W}B,\\
Var(\hat{m}(x)|\overline{X_1},\cdots,\overline{X_M}) &= \frac{\sigma^2}{n}e_1^T(\overline{X}^T\overline{W}\overline{X})^{-1}\overline{X}^T\overline{W}\overline{W}\overline{X}(\overline{X}^T\overline{W}\overline{X})^{-1}e_1,
\end{aligned}
\tag{10}
$$

where $e_1=(1,0)^T$. The variance term in (7) is obtained directly by (10). Then, we consider the bias term only. Because the support of K is $[-1,1]$ and $L_i/h\to 0$, we need only consider $|\overline{X_i}-x|<h$ and $|X_{ij}-x|\le|X_{ij}-\overline{X_i}|+|\overline{X_i}-x|\sim h$, and thus

$$
m(X_{ij})=m(x)+m'(x)(X_{ij}-x)+\frac{m''(x)}{2}(X_{ij}-x)^2(1+o_p(1)). \tag{11}
$$

By (11) and (10), we obtain

$$
E(\hat{m}(x)|\overline{X_1},\cdots,\overline{X_M})=m(x)+\frac{m''(x)}{2}\frac{1}{\overline{S_0}\,\overline{S_2}-\overline{S_1}^2}(\overline{S_2}\,\overline{V_0}-\overline{S_1}\,\overline{V_1})(1+o_p(1))
$$

where $\overline{V_l}=\sum_{i=1}^{M}\{\sum_{j=1}^{n}(X_{ij}-x)^2/n\}(\overline{X_i}-x)^l K_h(\overline{X_i}-x)/M$, for $l=0,1$. We can easily show that

$$
\begin{aligned}
\overline{V_l} &= \frac{1}{M}\sum_{i=1}^{M}K_h(\overline{X_i}-x)(\overline{X_i}-x)^l\frac{1}{n}\sum_{j=1}^{n}\{(X_{ij}-\overline{X_i})+(\overline{X_i}-x)\}^2\\
&= \overline{T_l}'+\overline{S}_{l+2},
\end{aligned}
$$

for $l=0,1$. This complete the proof of (7).

References

[1] Aitkin, M. and Longford, N. (1986). Statistical modeling issues in school effectiveness studies, *Journal of the Royal Statistical Society, Series A*, vol. 149, 1-43.

[2] Carroll, R.J. (1997). Surprising effects of measurement error on an aggregate data estimator, *Biometrika*, vol. 84, 231-234.

[3] Cleveland, W. (1979). Robust locally weighted regression and smoothing scatterplots, *Journal of the American Statistical Association*, vol. 74, 829-836.

[4] Fan, J. (1992). Design-adaptive nonparametric regression, *Journal of the American Statistical Association*, vol. 87, 998-1004.

[5] Fan, J. and Gijbels, I. (1996). *Local polynomial modelling and its application*: Chapman and Hall, London.

[6] Fan, J. and Marron, J.S. (1994). Fast implementations of nonparametric curve estimators, *Journal of Computational Graphical Statistics*, vol. 3, 35-56.

[7] Hall, P., Park, B.U. and Turlach, B.A. (1998). A note on design transformation and binning in nonparametric curve estimation, *Biometrika*, vol. 5, 469-476.

[8] Härdle, W. and Grund, B. (1991). Comment on "Choosing a kernel regression estimator" by C.K. Chu and J.S. Marron. *Statistical Science*, vol. 6, 421-425.

[9] Härdle, W. and Scott, D.W. (1992). Smoothing by weighted averaging of rounded points, *Computational Statistics*, vol. 7, 97-128.

[10] Müller, H.-G. (1987). Weighted local regression and kernel methods for nonparametric curve fitting, *Journal of the American Statistical Association*, vol. 82, 231-238.

[11] Prentice R.L. and Sheppard L. (1995). Aggregate data studies of disease risk factors, *Biometrika*, vol. 82, 113-125.

[12] Stone, C.J. (1977). Consistent nonparametric regression, *The Annals of Statistics*, vol. 5, 595-620.

[13] Wand, M.P. and Jones, M.C. (1996). *Kernel smoothing*: Chapman and Hall, London.